Int. J. on Recent Trends in Engineering and Technology, Vol. 8, No. 1, Jan 2013

The Effectiveness of 2D-3D Converters in Rendering Natural Water Phenomena

Syed A. Hussain[1], David F. McAllister[1]
[1] North Carolina State University, Raleigh NC, USA
Email: {sahussai, mcallist}@ncsu.edu

***Abstract*—Several commercially available conversion applications have been developed to generate 3D content from existing 2D images or videos. In this study, five 2D-3D converters are evaluated for their effectiveness in producing high quality 3D videos with scenery containing water phenomena. Such scenes are challenging to convert due to scene complexity including detail, scene dynamics, illumination, and reflective distortion. Comparisons are given using quantitative and subjective evaluations.**

***Index Terms*—2D-3D conversion, natural phenomena, stereo, imaging, parallax**

I. Introduction

3D display technologies have been widely deployed with success in movie industry, television, and in hand-held mobile devices. Release of numerous successful 3D movies in recent years has convinced many that stereoscopic 3D is here to stay. Meanwhile, hand-held devices, like smart phones and game consoles, have made it simpler to promote auto-stereoscopic displays with manufacturers looking at applying the technology to slightly larger displays of tablet computers. Despite this upsurge in viewing with stereoscopic 3D, there is a shortage of good quality 3D content. In contrast, content in 2D is abundant and readily available, thus creating opportunities to retrospectively convert existing 2D content into 3D.

The problem of converting 2D to 3D addresses the generation of left- and right-eye views with correct horizontal parallax from a given 2D view or video. This is a difficult problem to solve in realtime. In the movie industry, converting old movies to 3D is a meticulous, semi-automatic, and time consuming process. Many 3D television sets have a 2D-3D conversion mode, but the processing resources are limited, resulting in a poor quality visual experience. For computers, including tablets and hand-held devices, many fully automatic conversion algorithms have become available.

Producing realtime photo-realistic stereo in a scene with natural phenomena is particularly challenging due to scene complexity including detail, scene dynamics, illumination, and reflective distortion. References [1] and [2] describe methods for rendering stereo images of fire and gaseous phenomena in realtime. Among all natural phenomena, creating a convincing 3D impression of water is particularly difficult. The dynamics of water, its interaction with the environment, and light makes it a complex phenomenon to render in realtime. The alternative to modeling and rendering water phenomena is to use video of water scenery as an input to 2D-3D conversion software. Given accurate depth map estimation, such software applications may produce a stereo scene with water phenomena. However, in creating a depth map the 2D-3D video converters make many assumptions about the 3D scene and visual cues that are often not correct, resulting in conflicting 3D views. Also, the data available in the 2D input image of natural phenomena may not have enough information to give a look-around or immersive feel to the converted output image. It also does not solve hidden surface problems where changing the viewpoint changes the occlusion relationship between object in the scene. In this paper, we evaluate five commercially available 2D-3D video converters and study their effectiveness in adding depth to scenes containing water phenomena.

II. Overview

A. Recent Work

Existing 2D-3D conversion algorithms can be grouped in two categories: algorithms based on a single image and methods that require sequence of multiple images such as videos. Depth from a single still image can be extracted by employing monocular depth cues, such as linear perspective, shading, occlusion, relative size, and atmospheric scattering. Other techniques like blur analysis and image based rendering methods using bilateral symmetry also exist. McAllister [3] uses linear morphing between matching features to produce stereo output from a single image with bilateral symmetry, such as the human face.

For methods that require a sequence of multiple images, several heuristics exist to create depth information. These methods generate a depth map by segmenting the 2D image sequences, estimating depth by using one or combination of many visual cues, and augmenting the 2D images with depth to create left- and right-eye views. Reference [4] provides a detailed description the algorithms useful in computing dense or spare depth maps from multiple images of a scene either taken from similar vantage points or from a sequence of images acquired from a video. In another method, Hattori [5] describes realtime 2D-3D converter software that produces a 3D output viewable from different angles. To accomplish this, the author applies the horopter circle projection on the right-eye image. The horopter is the locus of points in space that fall on corresponding points in the two retinas when the left- and right-eye fixate on a given object in the scene. All points that lie on the horopter have no binocular disparity. In the absence of binocular disparity other depth cues such as linear perspective, shading, shadows, atmospheric scattering, occlusion, relative size, texture gradient, and color become

DOI: 01.IJRTET.8.1.56

more relevant. The author relates the parallax shift with pixel illumination assuming that brighter objects are closer to viewpoint while darker objects are in the background. This parallax shift method is used to create the left-eye view. The author further shows that the anaglyph output generated by this realtime 2D-3D converter produces less fatigue due to a decrease in retinal rivalry [6].

Other techniques apply machine learning algorithms and a classifier to automatically detect objects and key features in a given scene to estimate depth. Reference [7] describes one such algorithm where for each video frame a potential stereo match is determined by the classifier, which ensures that the proposed stereo pair meets certain geometric constraints for pleasant 3D viewing. In sequence of multiple images, the depth cues are also estimated by the presences of shadows, focus/defocus, disparity among two images, and motion parallax. There is extensive research on depth estimation in the context of 2D-3D conversion. Reference [8] provides an excellent overview of 2D-3D conversion techniques for 3D content generation.

In principle, depth can be recovered either from monocular or binocular depth cues. Conventional methods for depth estimation have relied on multiple images using stereo correspondence between two or more images to compute disparity. However, combining monocular and binocular cues together can give more accurate depth estimates [9]. The proliferation of depth estimation techniques has given rise to many practical software applications for 2D-3D conversion.

In this study, we compare five 2D-3D software applications that convert 2D video into stereoscopic 3D. These five applications are Arcsoft, Axara, DIANA-3D, Leawo, and Movavi. The goal of this study is to investigate the effectiveness of the 2D-3D converters in rendering natural water phenomena. We have not yet investigated conversion methods for other natural phenomena such as fire, smoke, fog, clouds, vegetation etc. The selection of these five applications is based on the method used, ease of use, and software availability. The DIANA-3D by Sea Phone [10] implements the method described above by Hattori. The Axara Media 2D -3D video converter software applies classifiers and automatic object detection in scenes to perform transformations from 2D to 3D video files [11]. The Arcsoft MediaConverter [12] uses proprietary 3D simulation technology to turn 2D pictures and movies into 3D format and is included in the study to compare how the algorithm compares to documented methods. The Leawo Video Converter [13] and Movavi Video Converter 3D [14] uses parallax shift and perspective to provide 2D to 3D video conversion support and are included to study how the two implementations compare.

III. Experiments And Results

A. Methodology

We are interested in measuring the quality of stereo output when a scene containing the natural phenomena of rain or water drops in motion and the effect of wind. In order to evaluate the output from the selected 2D-3D video converters we created a baseline synthetic video by using a 3D modeling tool. Additionally, we use a collection of downloadable stereoscopic videos of natural scenes acquired from an integrated twin lens camera system [15].

In these experiments, the quality of the stereo output is measured in two different ways. In the first method, we select two features in an input to the 2D-3D video converters such that one feature is closer to the viewer. Therefore, the correct output of the 2D-3D video converters has greater positive parallax between the left- and right-eye views in the feature that is farther from the camera. We compare the difference in the horizontal parallax between actual values and the values obtained by the output of the 2D-3D video converters. The second method to evaluate the quality of stereo output of the five 2D-3D video converters is based on subjective scoring by individuals who rate their overall visual experience. The results from both of these methods are described below.

B. Experiments

Baseline synthetic images are created to test 2D-3D video converters for commonly observed monoscopic depth clues such as linear perspective and occlusion. Fig. 1 shows one such test image emphasizing linear perspective. This 2D image of a 3D virtual scene is taken by two identical parallel camera models, one for each eye, giving a true 3D stereoscopic output. The scene consists of two identical spherical objects, representative of raindrops that are smaller than 2mm in size. The center of the sphere on the right is in the stereo window, while the sphere on the left is the same object further from the camera. The stereo window is a plane perpendicular to the viewer's line of slight on which the left- and right-eye views are projected. The stereo output image acquired by using the parallel camera models is the baseline output image, which is shown in Fig. 2. Notice that the left- and right-eye views of the sphere on the left shows greater positive parallax as it is placed away from the camera while the center of the sphere on the right has zero parallax and shows little disparity between the left- and right-eye views. This baseline output image is compared with the output of the five 2D-3D video converters. The horizontal parallax value is measured by identifying key features like edges or corners of an object in the left- and right-eye views. For test cases where the baseline image is from a stereo camera, a feature such as an edge or a corner of an object can be easily recognizable. The horizontal parallax of the selected feature from the baseline output image is the correct value. We measure the difference in this horizontal parallax for the same feature in the output of the 2D-3D video converters. The difference between the two values is the error in horizontal parallax introduced by the 2D-3D video converter. These experiments are repeated for depth implied by occlusion.

Figure 1. Baseline input image to test depth from linear perspective.

Int. J. on Recent Trends in Engineering and Technology, Vol. 8, No. 1, Jan 2013

Figure 2. Baseline output image to test depth from linear perspective shown as Left/Right/Left for parallel and cross stereo viewing.

For some methods objects at the bottom or center of the scene are assumed to be closer to the camera than objects at the top or near the edges. To test this scenario, a baseline image with objects appearing throughout the scene is used. Three videos consisting of a scene with water and wind effects are also used. The baseline output images used with actual parallax values are shown in the appendix. The following section describes the results of these experiments.

C. Results

The output of a 2D-3D video converter (Axara) to the input baseline image for linear perspective is shown in Fig. 3. It is expected that a feature closer to zero parallax would show little disparity. However, in the actual output of the 2D-3D video converter, the sphere closer to the camera, exhibits significant horizontal parallax. The comparison between positions of the selected features in the baseline image and the corresponding output from the five 2D-3D video converters is shown in Table I and Table II. The columns titled C-1 to C-6 correspond to the six different test cases used. The first three tests, from C-1 to C-3, are results from synthetic baseline input images while the results from C-4 to C-6 are from baseline images acquired from a stereoscopic camera. The values in Table I are horizontal parallax values for a selected feature that is closer to the camera. Table II shows the values for the objects that are farther from the camera. These values are measured in pixels. Notice that some values are negative. Negative values mean negative parallax. For example, in the baseline image, the sphere is positioned with the stereo window passing through the center; a portion of the sphere appears in front of the stereo window.

The test case C-1 corresponds to the depth from linear perspective. Arcsoft and Leawo are the only two converters that place the sphere with negative parallax. However, these values are in error when compared to the true value. The test case C-2 corresponds to depth due to variation of object placement in the scene. In this case, spheres are placed throughout the scene at various locations all with centers in the stereo window. We expect the output image to be at the same parallax unless the 2D-3D video converter is using scene placement to determine depth. Comparing C-2 with the same column in Table II should give the same values. Arcsoft is only 2D-3D video converter that exhibits different parallax values for objects placed at the bottom of the scene as opposed to the same object placed on the top.

Figure 3. Output image from a 2D-3D converter shown as Left/Right/Left for parallel and cross stereo viewing.

The C-3 test case includes occlusion. In the baseline image, the sphere farther from the camera is placed behind the sphere that is near the camera so it is partially occluded. The horizontal parallax value in column C-3 of the two tables for the baseline image confirms this fact. The remaining values in the column C-3 show that none of the 2D-3D video converters distinguished between the two spheres and the horizontal parallax values for the two spheres are the same.

TABLE I. PARALLAX (IN PIXELS) FOR OBJECTS CLOSER TO THE CAMERA

Converters	C-1	C-2	C-3	C-4	C-5	C-6
Baseline	-2	-2	-2	-12	-18	-14
Arcsoft	-6	-4	-4	-9	-6	-4
Axara	20	20	20	60	55	64
DIANA-3D	6	6	6	12	15	14
Leawo	-4	-4	-4	-17	-20	-20
Movavi	10	10	10	32	30	31

The test cases from C-4 to C-6 correspond to the videos of natural scenes taken from a stereoscopic camera. The horizontal parallax in the three baseline images is negative. The Arcsoft output for test cases C-5 and C-6 is visually conflicting as the feature farther from the camera has less horizontal parallax than feature closer to the camera. Axara, DIANA-3D, and Movavi outputs all have positive parallax and are therefore incorrect. Only the Leawo output exhibited negative parallax for all objects close to the camera. An important note from the data in Table I and Table II is that apart from Arcsoft, all other 2D-3D video converters showed no difference in horizontal parallax for features closer or farther from the camera, thus adding an equal amount of parallax to all objects. This simply gives a perception of the entire scene appearing behind or in front of the stereo window. The depth perception in these outputs is mainly due to monoscopic depth cues.

It is also noted that out of five 2D-3D video converters, four converters (Axara, DIANA-3D, Leawo, and Movavi) offer a user adjustable 3D depth setting. For the experiments this setting was set to the default value. The effect of changing the 3D depth setting results in either shifting all objects in a scene to appear behind or in front of the stereo window, thus adding either positive or negative parallax to the entire scene.

TABLE II. PARALLAX (IN PIXELS) FOR OBJECTS FARTHER FROM THE CAMERA

Converters	C-1	C-2	C-3	C-4	C-5	C-6
Baseline	26	-2	3	-8	8	-5
Arcsoft	-6	-8	-4	-4	-9	-8
Axara	20	20	20	60	55	64
DIANA-3D	6	6	6	12	7	12
Leawo	-4	-4	-4	-17	-20	-20
Movavi	10	10	10	32	30	31

We note that of the five 2D-3D video converters, DIANA-3D is the only converter that can convert a video in real-time. All other 2D-3D converters first uploaded a 2D video file before writing the converted 3D output.

DOI: 01.IJRTET.8.1.56

Int. J. on Recent Trends in Engineering and Technology, Vol. 8, No. 1, Jan 2013

From the data in Table I and Table II, a mean square error (MSE) value for each 2D-3D video converter for the feature closer and the feature farther from the camera is computed. For a given test case, the error is the difference between the parallax values of the baseline and the parallax value of the 2D-3D video converter. This error is squared and summed up for all test cases for that particular 2D-3D video converter. A mean value is calculated by dividing the squared sum values with the total number of test cases. Table III shows normalized mean squared error for the five 2D-3D converters. The data shows that Leawo had the least amount of error while Axara had the highest error, followed by Movavi. The errors in Arcsoft and DIANA-3D are close; Arcsoft performs better for closer objects while DIANA-3D has a smaller error for objects further from camera.

D. Subjective Scoring

Twenty five subjects participated in the subjective scoring. The experiment was performed by showing output of the five 2D-3D video converters and asking participants to rate their overall quality of visual experience. The rating is based on a five point scale where 1 is poor, 2 is marginal, 3 is average, 4 is good, and 5 is excellent. The input video is a synthetic rain scene shown in Fig. 4.

The participants expressed difficulty in observing depth in the scene. The rain streaks were difficult to observe as the resolution of the output anaglyphs is not high definition. The rain streaks are observed to be at the same depth even though the synthetic scene does plot some rain streaks closer to camera then the others. The results from these experiments are graphed in Fig. 5, which shows that Leawo produced the best results followed by DIANA-3D, Arcsoft, Movavi, and Axara. These results corroborate the normalized MSE values acquired from the previous experiments.

IV. Conclusions

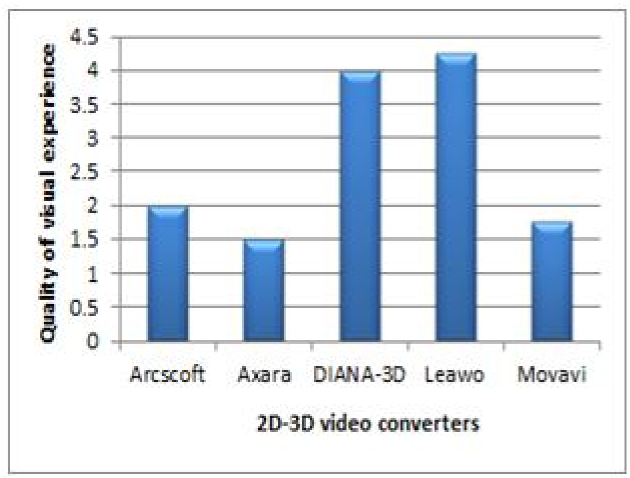
The experiments show the relative performance among five selected commercially available 2D-3D video converters. All six test cases were applied to measure horizontal parallax. Additionally, subjective scoring by twenty five participants measured the overall quality of visual experience. It is observed that the depth perception is mainly due to presence of strong monoscopic depth cues. The binocular disparity is equally applied to all objects in the scene, thus making the entire 2D image plane shift into or out of the screen. The 2D-3D video converters are making assumptions about the 3D scene that are often not correct, thus giving conflicting visual cues. The quality of the visual experience for scenes in this experiment is poor and we have developed other methods to enable realtime photo-realistic rendering of water phenomena. Preliminary results will be reported in [16].

Table III. Normalized MSE between Baseline and 2D-3D Converters

Converters	Near	Far
Arcsoft	0.015	0.115
Axara	1.000	1.000
DIANA-3D	0.146	0.094
Leawo	0.004	0.165
Movavi	0.371	0.309

Figure 4. Rain scene used as input to 2D-3D video converters for subjective experiments.

For future experiments, we propose to use an automatic feature detection algorithm and apply stereo matching between left- and right-eye views to determine the horizontal parallax instead of using manual measurements. This will increase the number of feature points to compare and enhance the test set and corresponding conclusions.



Quality of visual experience rating:
1 = Poor
2 = Marginal
3 = Average
4 = Good
5 = Excellent

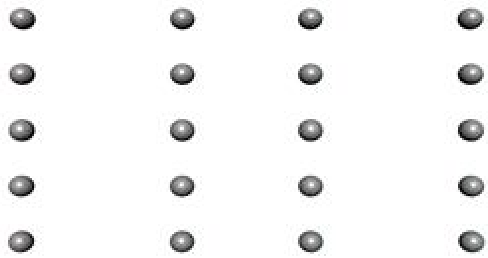
Figure 5. Results from subjective scoring

Appendix A Baseline Output Images

The baseline output image used to determine the true values for horizontal parallax for linear perspective is shown above in Fig. 2. The remaining five baseline images are shown in the appendix in Fig. 6. Two of these images are synthetic and the other three images are taken from videos of scenes with water and wind made from a stereoscopic camera.

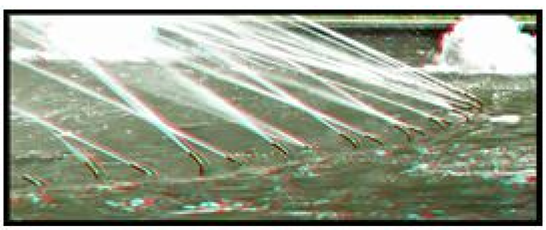
DOI: 01.IJRTET.8.1.56

Test case C-2



Test case C-3

Test case C-4



Test case C-5

Test case C-6

Figure 6. Baseline output images.

Acknowledgment

The authors wish to thank students of Wake Technical Community College, Raleigh, NC to volunteer their time in evaluating 2D-3D video converters and rate their overall visual experience for this study.

References

[1] Benjamin M. Rose and David F. McAllister, “Real-time Photorealistic Stereoscopic Rendering of Fire,” in *Proc. SPIE-IS&T Electronic Imaging*, vol. 6490, pp. 64901O- 64901O-13, February 2007.

[2] Tyler M. Johnson and David F. McAllister, “Real-time Stereo Imaging of Gaseous Phenomena,” in *Proc. SPIE-IS&T Electronic Imaging*, vol. 5664, pp. 92-103, March 2005.

[3] David F. McAllister, “Stereo Pairs from Linear Morphing,” in *Proc. SPIE-IS&T Electronic Imaging*, vol. 3295, pp. 46- 52, January 1998.

[4] Daniel Scharstein and Richard Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, *International Journal of Computer Vision*, vol. 47(1-3), pp. 7-42, 2002.

[5] Tomohiko Hattori “2D-3D Image Converter by Sea Phone” *The Journal of Three Dimensional Images*, vol. 23(2), pp. 36-39, 2009.

[6] Tomohiko Hattori, “Image Processing Concerning Individual Phenomena in Real Time 2D/3D Conversion Software DIANA for PCs”, *The Journal of Three Dimensional Images*, vol.25(2), pp. 14-18, 2011.

[7] Minwoo Park, Jiebo Luo, Andrew Gallagher, and Majid Rabbani, “Learning to Produce 3D Media from a Captured 2D Video”, *Proceedings of the 19th ACM international conference on Multimedia*, pp. 1557-1560, 2011.

[8] Liang Zhang, Carlos Vazquez, and Sebastian Knorr, “3DTV Content Creation: Automatic 2D-to-3D Video Conversion,” *IEEE Transactions on Broadcasting*, vol. 57(2), pp. 372-383, June 2011.

[9] Ashutosh Saxena, Jamie Schulte, and Andrew Y. Ng, “Depth Estimation using Monocular and Stereo Cues,” in *Proc. of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, Morgan Kaufmann Publishers Inc., pp. 2197-2203, San Francisco, CA, USA, 2007.

[10] SeaPhone Co., Ltd., “Diana Is a Real Time 2D/3D Image Converter for DVD,” http://shop.vrtc.co.jp/index.php?act=shop_en_goods.

[11] AxaraMedia Ltd., “2D to 3D Video Converter,” http://www.axaramedia.com/VideoSolutions/2dto3dconverter.

[12] ArcSoft Inc., “2D to 3D MediaConverter version 8,” http://www.arcsoft.com/mediaconverter.

[13] Leawo Software Co., Ltd., “Video Converter,” http://www.leawo.com/leawo-video-converter.

[14] Movavi, “Video Converter 3D,” http://www.movavi.com/videoconverter3d.

[15] E. Cheng, P. Burton, J. Burton, A. Joseski, and I. Burnett, “RMIT3DV: Pre-announcement of a Creative Commons Uncompressed HD 3D Video Database,” in *Quality of Multimedia Experience (QoMEX)*, Fourth International Workshop, pp. 212-217, July 2012.

[16] Syed A. Hussain and David F. McAllister, “Stereo rendering of rain in realtime,” *in Proc. SPIE-IS&T Electronic Imaging*, in press.

DOI: 01.IJRTET.8.1.56